Preparing to Run on Pleiades' Sandy Bridge Nodes

Category: Tips & Tricks

Overview of Sandy Bridge Compute Nodes

Pleiades has 24 Sandy Bridge racks, each containing 72 nodes. Each of the 24 nodes contains 2 eight-core E5-2670 processor chips and has 32 GB of memory.

The E5-2670 processor speed is 2.6 GHz when Turbo Boost is OFF, and can reach 3.3 GHz when Turbo Boost is ON. When Turbo Boost is enabled, idle cores are turned off and and power is channeled to the active cores, making them more efficient. The net effect is that the active cores perform above their clock speed (that is, overclocked).

The Sandy Bridge nodes are connected to the Pleiades InfiniBand (ib0 and ib1) network via the 4-link fourteen data rate (4x FDR) devices and switches for inter-node communication.

The Lustre filesystems, /nobackupp[*1-6*], are accessible from the Sandy Bridge nodes.

Compiling Your Code For Sandy Bridge Nodes

One important feature of the Sandy Bridge processor is its use of the Advanced Vector Extensions (AVX), a set of instructions for doing Single Instruction Multiple Data (SIMD) operations on Intel architecture processors. AVX uses 256-bit floating point registers, which are twice as wide as the 128-bit registers used in the Harpertown, Nehalem-EP and Westmere processors. With two floating-point functional units and 256-bit registers (which can hold four double-precision floating-point values or eight single precision floating point values), a code with well-vectorized loops can achieve a maximum of either eight double-precision floating-point operations (FLOPs) per cycle, per core or 16 single-precision FLOPs per cycle, per core.

To take advantage of AVX, we recommend that you recompile your code with an Intel version 12 compiler (for example, **`comp-intel/2011.7.256`**) on Pleiades, using either of the following compiler flags:

- **`-O2 (or -O3) -xAVX`**

  for an executable that only runs on Sandy Bridge
- **`-O2 (or -O3) -axAVX -xSSE4.1`**

for an executable that can run on all four Pleiades processor types

You may also add the compiler options **-ip** or **-ipo**, which allow the compiler to look for ways to better optimize and/or vectorize your code.

To get a report on how well your code is vectorized, add the compiler flag **-vec-report2**.

To compare the performance differences between using AVX and not using AVX, we recommend that you create separate executables, one with **-xAVX** or **-axAVX -xSSE4.1** and another one without them. If you do not notice much performance improvement using these flags, then your code does not benefit from AVX.

However, this does not mean that your code will not run faster on Sandy Bridge nodes than on Harpertown, Nehalem-EP or Westmere nodes, because your code can still benefit from other Sandy Bridge hardware improvements, such as larger L3 cache, higher memory bandwidth, and faster interconnects.

If you have an MPI code that uses the SGI MPT library, you should use the module **mpi-sgi/mpt.2.06a67**. This is because FDR is supported in MPT 2.06, but not in earlier versions (mpt.1.25, mpt.1.26, mpt.2.01, and all mpt.2.04 modules).

TIP: It is important to check the correctness of your runs on Sandy Bridge before production work.

Running PBS Jobs on Sandy Bridge Nodes

Of the 24 Sandy Bridge racks, 22 will be used for production runs via the **normal**, **debug**, and **long** queues. The remaining two racks will be used for development work via the **devel** queue.

To request Sandy Bridge nodes, use **:model=san** in your PBS script:

```
#PBS -l select=xx:ncpus=yy:model=san
```

To request the **devel** queue, use either of the following methods:

In your PBS script, add:

```
#PBS -q devel
```

In your **qsub** command line, use:

```
pfe% qsub -q devel your_pbs_script
```

Since there are 16 cores per Sandy Bridge node compared to 12 cores per Westmere node and 8 cores per Nehalem-EP or Harpertown node, a PBS job running with a fixed number of processes or threads should use fewer Sandy Bridge nodes than the other three processor types.

For example, if you used to run a 48-process job with 4 Westmere nodes, or 6 Nehalem-EP or Harpertown nodes, you should request 3 Sandy Bridge nodes instead.

For Westmere
```
#PBS -lselect=4:ncpus=12:mpiprocs=12:model=wes
```

For Nehalem
```
#PBS -lselect=6:ncpus=8:mpiprocs=8:model=neh
```

For Harpertown
```
#PBS -lselect=6:ncpus=8:mpiprocs=8:model=har
```

For Sandy Bridge
```
#PBS -lselect=3:ncpus=16:mpiprocs=16:model=san
```

Sample PBS Script For Sandy Bridge

```
#PBS -lselect=3:ncpus=16:mpiprocs=16:model=san
#PBS -q devel

module load comp-intel/2011.7.256 mpi-sgi/mpt.2.06a67

cd $PBS_O_WORKDIR

mpiexec -np 48 ./a.out
```

For more information about Sandy Bridge nodes, see:

- Pleiades architecture overview
- Sandy Bridge Processors

Article ID: 322
Last updated: 18 Dec, 2012
Tips & Tricks -> Preparing to Run on Pleiades' Sandy Bridge Nodes
http://www.nas.nasa.gov/hecc/support/kb/entry/322/?ajax=1